

## Learn Open Source Database Tools from Stanford for Free



I recently finished Stanford's excellent free on-line course [Introduction to Databases](#) with [Jennifer Widom](#). The course is a broad survey of database technology including [XML](#), [Relational Database Management Systems \(RDBMS\)](#) from many angles ([SQL](#) forms the centerpiece of the course), [OLAP \(OnLine Analytical Processing\)](#) and [NoSQL](#).

I was very impressed with the breadth of Widom's approach to the subject: it was a major reason I decided to spend time on the course. Another strength is its nuts-n-bolts approach: some theoretical topics are covered but for the most part this is a course for practitioners. Finally, I particularly appreciated the extensive use of FOSS (Free and Open Source Software) in the course.

Why study databases? I will merely say that data is a core tool pervading the information resources of modern civilization. [Databases](#) are where data is housed. For example, the data constituting this blog is stored in a database and the same can be said for much (if not all) of the Internet. Databases are a profoundly vital, big picture subject.

Widom's course is still open for enrollment in "archival mode" meaning you can watch the videos, work through the exercises, quizzes and exams, and track your progress, but the deadlines have expired and no more "Statements of Accomplishment" will be awarded (at least until the course is offered again). To complete the simple enrollment go to [db-class.org](#) and start learning about databases with FOSS today!

Although the course is broadly useful for anyone wanting to learn the basics of databases from a broad perspective, I found it to be particularly good for learning about the FOSS tools that can support database systems. So let's start there.

### **FOSS Tools Covered**

For traditional RDBMS, the course uses [PostgreSQL](#), [SQLite](#), and [MySQL](#). Widom mentions some limitations of each database (DB) in regards to the SQL (Structured Query Language) standard including important distinctions about using each system with [triggers](#), [transactions](#) (PostgreSQL has the best support), [Views](#) (MySQL uses [updatable views](#) whereas PostgreSQL & SQLite use triggers to modify views), recursion (only PostgreSQL and only in newer versions), and OLAP (only MySQL supports [with rollup](#)).

On the XML side, I learned [xmllint](#) and [SAXON](#) for XML validation, querying and transformation. The course covers the basics of [DTDs](#), [XML Schema](#), [XPath](#), [XQuery](#) and [XSLT](#). I used xmllint, saxonb-xquery, and saxonb-xslt to work through the exercises (the searchable Q&A forum provides usage details).

Finally, for NoSQL there are two videos which survey the state of the art. There is some depth on [the Map Reduce framework](#) provided by [Hadoop](#). Several other FOSS systems are briefly discussed: [Cassandra](#), [Voldemort](#), [CouchDB](#), [MongoDB](#), and more. The NoSQL portion of the

course is a good overview of the technology, but there are no exercises and hence little depth of a concrete nature.

From a FOSS perspective, the course is exquisite: FOSS utilities were front and center for the duration and some guidance in using these tools is provided. Help was readily available: I answered a few questions in the Q&A forum to help people overcome hurdles and I used its search feature to overcome some of my own. In sum, studying this course will give you the lay of the land for FOSS database technology including some advice about the limitations and strengths of its best database tools.

## Stanford's innovative platform for free on-line video courses

I am a big fan of so-called [Open Educational Resources \(OER\)](#) including free on-line video courses. Stanford's Databases course is the 12th I've completed, but only the second in which I did a "deep dive" by reinforcing learning with exercises, quizzes and exams. In general, I use OER video courses as [edutainment](#) as I usually find the extra work too time-consuming: my goal is to broadly understand how the world works, not to build expertise in every subject I study! So, conceptually, I prefer the traditional form of video courses pioneered by [MIT's OpenCourseWare](#) which in contrast with Stanford's new approach might be called **archived courses**. Archived courses make the material available without (m)any social tools. So, working through the materials in traditional OER courses usually requires extra self-discipline and commitment (unless you just watch the videos for fun as I often do).

[Stanford's OER system, online at coursera.org](#), builds on the basic idea of OER video courses by adding deadlines, interactive feedback from automatically evaluated work, and some, including the Databases course, offer the ability to earn a "Statement of Accomplishment" for demonstrating basic proficiency. It is precisely these social enhancements that makes Stanford's initiative so noteworthy. Together these social tools provide a shared experience with a clear set of tasks for a cohort of students working through the course at the same time.

The extra interactivity and the focus of deadlines give the Stanford approach to OER a special excitement and sense of goal accomplishment which is absent in archived courses. Even though I prefer the archived courses whose videos can be more entertaining than Stanford's tutorial-focused approach, I have to admit I was enthralled by the deadlines: they kept me focused. It should be emphasized that Stanford's courses like the more traditional OER archival courses can be pursued at a pace that suits your time and interest: there's no imperative to follow the deadlines or earn kudos for accomplishments.

## How I used the course materials

Although I have been using databases professionally for many years, I had not read nor studied the subject in any depth previously. I decided to take this opportunity for a deep dive. Widom's Databases course includes a simple enrollment process, tutorial-style videos (for download or in a browser with Flash support), automatically graded exercises, quizzes, and exams each with hard deadlines, a Course Materials section with many goodies, Optional Exercises, a FAQ, a Q&A forum, and a "Statement of Accomplishment" from the instructor if you completed a substantial portion of the coursework by the deadline (6,513 of the 91,000 students enrolled in the Fall 2011 Databases course earned one; [here is mine](#)).

First, I watched each video twice taking detailed handwritten notes on the second viewing (22 pages worth!). I then checked the Flash version of the videos which often included inline questions that were very useful (Stanford's Flash video viewer is the best I've seen: it even supports speeding up the video by 1.2 or 1.5 times while automatically adjusting the pitch!).

Then I worked through the quizzes and exercises. One nice feature of both was that you could attempt them many times. Different variants were provided to many of the quiz questions to make it harder to apply a blind trial and error approach and you can continue to work on them after getting them all correct (which might be useful as a way to practice for the exams or to see if you remember anything of the course when you check back in 1 or 10 years). I found the quizzes and exams to be very challenging and not so rewarding. Of course some mastery is required and acquired from the quizzes and for other learning styles they may prove more valuable than they were for me: judge for yourself.

The course included many supplementary exercises to provide extra practice. I used them for Relational Algebra and they were very helpful. However due to time constraints, I was unable to use them further. Most of my time was spent working through the exercises. I did all the exercises in an xterm window running sqlite, psql, xmllint, saxonb-xquery, or saxonb-xslt and pasting the results into the query workbench. This allowed me to really experience the FOSS tools "in the wild" which gave me a strong sense of their ins and outs. For me the interactive exercises were fantastic: they really helped me learn the material by directly engaging my problem solving faculty. They were like a real project with deadlines! Although I occasionally got ruffled with some of the difficult ones, they were engaging and fun!

Although the course web site is very simple and well-designed, it was still possible to have difficulty finding some of the gems provided for the students. For example, it took me awhile to find the code used in the demos (which was extremely useful by the way): the Course Materials section of the site has all the goodies you need but you have to mouse over the icons to see treasures that appear hidden at first (remember to right click to download). Also look carefully at the prescripts or postscripts affixed to some sections: more treasures.

The Q&A forum was helpful for finding things that were not at first apparent. A couple of times I scoured the Internet or Wikipedia looking for other angles on the material to understand a point I was struggling with. All work for the class is "open book", so I only prepared for the exams by simply reviewing my notes.

## Advice for students

I recommend taking the course now even though the deadlines have expired. Feel free to skip any part of the course that your interests, time constraints, and patience warrant. If the course is ever offered again, you will already know much of the material which should help you earn a "Statement of Accomplishment". If not, you will better understand a broadly useful, important and interesting subject.

If you want to do a "lite" version of the course, I recommend skipping the quizzes and exams. In addition, many of the topics can be skipped if you are short on time or find them uninteresting. To her credit Jennifer Widom recommended as much in her [screen side chats](#) which provided a wonderful human dimension to the course. Although some of the material is cumulative, there are several parts of the course for which skipping is a real option. For example, the Relational Algebra

(I thoroughly enjoyed doing those exercises!!!) and the Relational Design Theory topics are, I think, less important especially if you just want to acquire basic [DBA](#) skills. I deal with enough XML, that I found that part of the course extremely useful, but I can imagine someone who just needs SQL might skip those parts. This is a free course: you can be creative in how you use the materials so that you get what you want out of it!

I recommend doing the exercises associated with each topic (some did not have exercises). Some of them are quite challenging and some took me quite a bit of time. If necessary, skip some of the harder ones. If time is really pressing, just watch the videos that particularly interest you: remember this is a free resource: you can tailor your work on the course in whatever way suits your interests and time.

I did not buy nor borrow a textbook for the course (I was very impressed that Widom prepared reading assignments for four separate texts in the Course Materials sections of the site. Wow, that must have been a lot of work!). Having been through the course, I think a text is unnecessary for most students. You may find a few topics that are hard for you or for which the videos were insufficient to master the material. Since textbooks are more comprehensive and more detailed, they could help fill in the gaps. In particular diligent students may want a text. I prefer to learn iteratively, that is, I would prefer a shallow course today and another later that goes in more depth (I might even prefer to take two lite courses to build my mastery of a subject by degrees). But if you want a more complete experience now, then by all means get a textbook and dig in!

## Other OER Database Resources

In addition to the OER resources below, I found occasion to reference [Widom's course site for Stanford students in the allied CS145 Introduction to Databases](#) and her colleague [Jeff Ullman's offering of CS145 from Autumn 2002](#). [MIT OCW has 6.814/6.830 on Database Systems](#) which looks a bit too advanced for an introductory course and [MIT OCW has 1.264J/ESD.264J Database, Internet, and Systems Integration Technologies](#) which I found useful to supplement Widom's course (especially for Relational Design Theory). Finally, the [Indian Institutes of Technology \(IIT\)](#) has a complete video course (with 43 videos totaling 40 hours and 17 minutes) on [Database Management Systems](#) or watch its [YouTube playlist](#) (I did not look at these videos, but I've seen other IIT material and they are usually very informative and accurate).

## Conclusion

[Stanford's exciting new system for on-line courses](#) is remarkable in its use of social tools to engage students. This is a boon to the [OER](#) movement! In today's rapidly changing world, refreshing and expanding one's skills is essential to apprehending the needs and opportunities that abound if you are curious enough!

I'm hooked! Although I still love and will continue to use archived courses, I will be checking to see if Stanford has any courses of interest regularly from now on! I've signed up for several of Stanford's offerings (there are 16 of them!) for Winter term with the intention of "dropping" or doing a "shallow dive" (maybe just watch a few videos and do some exercises as interest permits). But I am eying the [Model Thinking course](#) for a possible deep dive. To see the full list of offerings go to [Class Central: Summary of Stanford's online course offerings](#) and plan your learning for the Winter term which starts next Monday, January 23rd!

Stanford's [Fall 2011 edition of Introduction to Databases](#) was a great course! Kudos to Jennifer Widom and everyone at Stanford who made this possible.

If you haven't started or finished it yet, head over to [db-class.org](#) now and get to it! It is one of the best resources on the Internet for learning about databases. Moreover, it includes the special benefit of covering a broad range of important FOSS database tools.

### **Other Reviews of Widom's Databases Course**

- [Franklin Chen's excellent review of Widom's Introduction to Databases course](#)
- [John Johnson's statistician's view of Stanford's open Introduction to Databases class](#)
- [Rio Akasaka's reflection on Stanford's AI-class, DB-class](#)

*Posted by CJ Fearnley in FOSS Community, 5 comments*